Optimal Feature Selection for Artifact Classification in EEG Time Series

Vernon Lawhern[1], W. David Hairston[2], and Kay Robbins[1]

[1] Department of Computer Science, University of Texas-San Antonio, San Antonio, TX 78249 USA
[2] Human Research and Engineering Directorate, Army Research Laboratory, Aberdeen Proving Ground, MD 21005 USA
{Vernon.Lawhern,Kay.Robbins}@utsa.edu,
william.d.hairston4.civ@mail.mil

Abstract. Identifying artifacts or non-brain electrical signals in EEG time series is often a necessary but time-consuming preprocessing step, as many EEG analysis techniques require that the data be artifact free. Because of this, reliable and accurate techniques for automated artifact detection are desirable in practice. Previous research has shown that coefficients obtained from autoregressive (AR) models can be used as feature vectors to classify among several different artifact conditions found in EEG. However, a statistical method for identifying significant AR features has not been presented. In this work we propose a method for determining the optimal AR features that is based on penalized multinomial regression. Our results indicate that the size of the feature vector can be greatly reduced with minimal loss to classification accuracy. The features selected by this algorithm localize to specific channels and suggests a possible BCI implementation with increased computational efficiency than with using all available channels. We also show that the significant AR features produced by this approach correlate to known brain physiological properties.

Keywords: Autoregressive (AR) model, Artifacts, Electroencephalography, classification, feature selection, multinomial regression, penalized regression, machine learning.

1 Introduction

In current EEG, the signal of interest is easily confounded by other biological sources of voltage, often stemming from muscle (EMG) or eye (EOG) movements. Great care is taken in laboratory settings to limit sources of artifacts, such as by having subjects limit any unnecessary movements or actions during the experiment, as these activities may confound the EEG activities of interest. After completing an experiment, researchers still must remove artifacts in EEG signals to obtain a “clean” signal that can be further analyzed. This process often requires manual identification of artifact-contaminated EEG, generally conducted by a panel of experts, which can be tedious and time-consuming, especially for large amounts of data. New applications of EEG

D.D. Schmorrow and C.M. Fidopiastis (Eds.): AC/HCII 2013, LNAI 8027, pp. 326–334, 2013.

are being performed in more complex and realistic environments, where controlling the effects of artifacts is not feasible, such as in the detection of fatigue while driving [1]. Similarly, brain-computer interfaces (BCIs) are being developed for individuals who may have physical disabilities and as a means to improve performance in healthy individuals [2]. In these scenarios, traditional labor-intensive off-line analyses that require extensive computation to remove artifacts are not feasible. Thus, extending applications of EEG to more realistic scenarios will require automated artifact detection methods that are robust to both inter-subject and intra-subject variations.

A common approach for the analysis of EEG signals is autoregressive (AR) modeling. Autoregressive models are linear models that relate signals to their past values. The coefficients of these models can characterize signal properties. Single-channel AR models relate signals to their own values, while multivariate models can model relationships between simultaneously-recorded time series. Useful characteristics of time series can be derived such as ordinary, partial or directed coherence [3, 4] and the direct transfer function (DTF) [5]. AR models are attractive representations in that they are compact and computationally efficient.

One important feature of AR models is that the coefficients are invariant to scaling changes in the data, making AR approaches valuable in EEG analyses. AR modeling has been extensively used in EEG data analysis for feature extraction and classification tasks [6], detection and classification of cardiac arrhythmias [7], and analysis of epilepsy data [8]. AR models have also been used for detecting artifacts in EEG signals. For example, Van de Velde et al [9] used features such as the slope, signal variance and AR model coefficients to classify EEG segments into three artifact categories: None, Moderate and Severe.

Our recent work [10] has shown that AR coefficients can be used alone to classify type-specific artifacts such as eye blinks and jaw movements. This method uses AR coefficients together with a support vector machine (SVM) classifier to distinguish among 8 different artifact conditions. While this is already a relatively efficient method, the high degree of correlation in the signals from neighboring channels and the close relationship of their resulting AR features exhibit a high degree of redundancy if all channels of a high-density cap are included. This suggests substantial room for streamlining the computation and very likely the hardware necessary for data acquisition. Channel elimination requires reliable methods for down-selecting channels, and the high degree of correlation among the features may make traditional feature selection techniques such as AIC (Akaike information criterion) or BIC (Bayesian information criterion) unreliable. There may also be situations where there are many more parameters than samples (the $p >> N$ case), making this an ill-posed problem which cannot be solved using traditional methods. In addition to reasons of analysis, using fewer features has advantages for implementation in natural environments using portable EEG headsets, which usually have many fewer channels than high-density laboratory models. In this environment, processing must be done online and not all channels may be in full contact. Therefore it is valuable to investigate methods that can be used to select only the most important features for EEG signal classification and to understand more clearly how information from different channel loci contribute to classification of different artifact types.

In this paper, we propose a method for determining significant signal features for artifact classification based on regularized multinomial regression. Multinomial regression is an extension of logistic regression, where more than two response classes are present. We use the artifact classes found in [10] as the response levels while using the AR coefficients from EEG channels as the covariates in the model. Since the AR coefficients exhibit a high degree of multi-variable co-linearity, we use an elastic net penalization [11] of the standard maximum likelihood solution to determine the optimal features. This approach has been used successfully in situations where there are many more parameters than samples ($p >> N$) such as in microarray gene expression data and text classification [12]. The high degree of co-linearity can make the matrix inversions needed for standard maximum likelihood unreliable and inaccurate. Our results indicate that a significant reduction in the feature set size is possible without loss in classification accuracy.

2 Experimental Methods

2.1 Experimental Setup

The data used in this study was recorded using a 64-channel Biosemi ActiveTwo System and analyzed in a previous study [10]. A brief summary is given here. A total of seven participants performed a block of artifact-inducing facial and head movements. All provided consent prior to participating, and methods were approved as required by U.S. Army human use regulations [13, 14]. The seven movements included (abbreviations follow): clenching the jaw (JC); moving the jaw vertically (JM); blinking both eyes (EB); moving eyes leftward, then back to center (EL); moving eyes upwards, then back to center (EU); raising and lowering eyebrows (ME); and rotating head side-to-side (as in looking leftward), (RH). All movements were performed sitting in front of a PC screen. The participants were instructed to perform each type of movement 20 times in concert with a consistently occurring tone. A baseline dataset was also recorded for each participant. Participants were told to look straight at the computer screen and to not move excessively in order to minimize muscle artifacts. We extracted 20 epochs of each artifact condition, plus 20 artifact-free epochs from the baseline condition. Our total dataset consisted of 160 epochs, 20 for each of 8 conditions for each of seven participants (see [10] for more details).

3 Statistical Methods

3.1 Autoregressive Models

We use autoregressive (AR) model coefficients as features for artifact classification in EEG. Given a zero mean time series $z_t, t = 1, \ldots, n$, an AR model of order p can be written as:

$$z_t = \sum_{i=1}^{p} A_i z_{t-i} + \epsilon_t \tag{1}$$

where $A_{i,}, i = 1, \ldots, p$ are the AR model coefficients, and $\epsilon_t \sim N(0, \sigma^2)$. The AR model estimates the signal characteristics by modeling the signal compared to the signal in the past p time points. In our analysis EEG channels are modeled individually using a second order AR model, and the AR coefficients are concatenated across channels to form the feature vector used for classification, resulting in a 128-dimensional feature vector. We use the Burg method for fitting the AR coefficients [15].

3.2 Multinomial Regression with Elastic Net Penalization

We treat the classification of artifact signals as a multinomial regression problem, where the artifact classes are the response levels, and the covariates are the AR coefficient features. Let $Y \in \mathbb{R}$ be the response variable (consisting of artifact labels) and $X \in \mathbb{R}^{\mathrm{C}}$ be the vector of AR coefficients (C = 128). Using the notation from [11], the multinomial regression model for the response variable G, having K > 2 levels, is:

$$\Pr(G = l|X) = \frac{e^{\beta_{0l} + X^T \beta_l}}{\sum_{k=1}^{K} e^{\beta_{0k} + X^T \beta_k}} \tag{2}$$

where $l = 1, \ldots, K$. We fit this model using regularized multinomial maximum likelihood. Let $p_l(x_i) = \Pr(G = l|x)$ and let $g_i \in \{1,2, \ldots, K\}$ be the i^{th} response. The penalized log-likelihood is:

$$\max_{\{\beta_{0l}, \beta_l\}_1^K \in \mathbb{R}^{K(p+1)}} \left[\frac{1}{N} \sum_{i=1}^{N} \log_{p_{g_i}}(x_i) - \lambda \sum_{l=1}^{K} P_\alpha(\beta_l)\right] \tag{3}$$

where λ is the penalty coefficient and:

$$P_\alpha(\beta_l) = \sum_{j=1}^{C} \left[\frac{1}{2}(1 - \alpha)\beta_{lj}^2 + \alpha|\beta_{lj}|)\right] \tag{4}$$

is the *elastic net penalty* [11]. This penalty reduces to the ridge regression penalty when $\alpha = 0$ (the standard l_2 penalty) and the Lasso penalty when $\alpha = 1$ (the standard l_1 penalty). The Lasso penalty is a sparse penalty that forces many of the coefficients to be 0, with a small subset to be nonzero, while the ridge regression penalty shrinks the coefficients of highly correlated variables relative to each other. The parameter α controls the degree of homogeneity among the two penalties. Setting $\alpha = 1 - \epsilon$ for some small ϵ produces a sparse solution similar to Lasso as well as removing irregular behavior caused by a high degree of co-linearity among the covariates. In our analysis we set $\alpha = .99$ as we seek a sparse solution that is robust to high correlations among covariates. We use the GLMNET toolbox for MATLAB [11] to solve for the coefficients. The optimal λ is found by using a grid search and maximizing the percentage of explained deviance (see [11] for more details).

3.3 Bootstrap Model Validation

To verify the significance of the model parameters, we randomly partitioned our data into two sets, a training (60%) and testing (40%) set. The training set is used to fit the regularized multinomial model, while the testing set is used to validate the accuracy of the classification. We used B = 100 bootstrap samples and calculated the average accuracy across all the samples. Note that the significant covariates may change at each bootstrap iteration; therefore, in a separate analysis, the covariates that were significant in at least 75% of the bootstrap iterations were extracted and a ridge regression model ($\alpha = 0$) was used on only these covariates. A ridge regression model was used as high degree of co-linearity may still exist among these covariates.

4 Results

The results of our classification study are shown in Table 1. The first row within each subject grouping denotes the classification accuracies when using all available parameters in the data and using the radial basis function support vector machine (RBF-SVM) that was used in [10] for artifact classification. The results from this classification are taken as the baseline performance, which we compare our current methods against. The average classification accuracy over all subjects is 95.8% +/- 2%. The second row denotes the classification accuracy from the elastic net penalty for the multinomial regression. The average performance in this case is not significantly different than using the full feature vector with SVM (94.7% +/- 2.4%) while using significantly fewer parameters in the model (40.3). This result indicates that the AR feature vector is highly redundant and in fact the majority of features are not necessary to obtain the same classification accuracy. When using only the parameters that appeared in at least 75% of the bootstrap iterations (third row within subject), we see a slight reduction in accuracy of ~4-5%. A Kruskal Wallis ANOVA revealed only minimal evidence of a significant difference in the three classification probabilities (χ^2 = 7.48, $p < .03$). Note that subject 7 saw no decrease in overall performance between the two models, while subjects 3 and 6 saw minimal reduction (3% or less).

Figure 1 shows a channel plot of significant channels for all of the subjects in the analysis. The first plot (top left) denotes the standard configuration of the 64-channel Biosemi System (see Materials and Methods). Channels in red indicate that at least one of the two AR(2) coefficients was significant in at least 75% of bootstrap samples, while channels in blue indicate both the AR(2) coefficients were significant in at least 75% of bootstrap samples. We see that there is some degree of consistency across subjects, with channels located frontally significant, while a few channels around the edge of the cap are also consistently contributing to the discrimination.

Figure 2 shows the classification performance for different criterion percentage values of the bootstrap models. The x-axis value at 0 denotes the classification percentage using the full feature vector (128 parameters) similar to the SVM-only classifier as in [10]. The bootstrap percentage value at 20 indicates that we use the parameters that occur in at least 20% of bootstrap models to build the multinomial regression. The two y-axes denote the percentage of the total number of parameters used in the model (blue, left side, which varies by percentage criterion) and the resulting overall classification percentage (green, right side). For example, at the bootstrap percentage value of 20% (meaning parameters had to appear in at least 20% of the bootstrap models to be included for analysis), about 40% of the parameters were used (~54 parameters) while achieving a classification percentage of ~93%. While there is a dramatic drop in the percent of parameters remaining in the model, which tapers to a slower decline, we simultaneously see that the accuracy curve (green) remains fairly flat until after the 80% bootstrap percentage value, where a noticeable reduction (to about 83%) occurs.

Table 1. Classification percentages for the elastic net regression models for classifying artifact conditions based on the average of 100 bootstrap models. Values in parentheses denote one standard deviation of the classification percentage. The first row within each subject denotes the average classification probabilities using all available parameters and using the SVM for classification. The second row denotes the average classification probability using elastic net penalization method, while the third row denotes the average classification probability only using parameters that were significant in >75% of the bootstrap models and using ridge regression to fit the multinomial model. The *P* column in the first row of each subject denotes the average number of significant parameters. Mean = average accuracy for all movements, JC = Jaw Clench, JM = Jaw Movement, EB = Eye Blink, EL = Eye Left Movement, EU = Eye Up Movement, ME = Move Eyebrows, RH = Rotate Head.

Subj	*P*	*Mean*	*JC*	*JM*	*EB*	*EL*	*EU*	*ME*	*RH*	*None*
1	128	96(1.8)	99(2.8)	99(2.8)	87(10.3)	93(6.3)	94(7.5)	100(0)	98(4.5)	96(7.1)
	38.7	95(2.4)	99 (1.2)	95(6.1)	89(10.9)	90(10.0)	98(5.1)	100(0)	97(5.7)	92(10.7)
	26	91(3.6)	94(10.2)	88(10.7)	91(7.9)	86(10.5)	97(5.8)	92(10.3)	93(9.8)	85(13.8)
2	128	93(2.4)	99(2.8)	92(7.3)	100(0)	97(5.5)	86(10.6)	89(7.3)	88(10.3)	89(12.3)
	49.7	90(3.3)	99 (3.2)	84(13.8)	98(3.7)	92(8.3)	90(10.9)	89(8.5)	75(13.3)	84(12.3)
	24	86(3.6)	99(2.1)	77(14.4)	98(4.3)	90(11.8)	81(14.5)	86(10.3)	84(13.3)	75(14.7)
3	128	97(2.3)	100(0)	89(10.1)	100(0)	100(0)	92(8.3)	100(0)	97(6.8)	96(6.1)
	35	98(1.6)	95(7.3)	94(9.4)	99(2.1)	100(0)	100(0)	99(1.2)	96(5.8)	100(0)
	19	95(2.1)	99(2.1)	90(9.4)	99(2.4)	98(5.2)	84(8.6)	98(4.2)	95(6.4)	94(6.5)
4	128	94(3.6)	100(0)	99(2.8)	98(4.5)	88(14.5)	85(16.5)	99(2.8)	94(11.8)	81(11.1)
	37.9	94(2.8)	99 (1.2)	100(0)	96(5.7)	90(10.3)	91(9.8)	98(4.7)	87(10.2)	91(10.1)
	20	89(3.6)	97(7.8)	99(2.4)	96(5.5)	83(12.1)	87(11.3)	91(9.7)	78(13.6)	79(14.1)
5	128	97(2.1)	100(0)	100(0)	99(2.8)	84(10.8)	93(10.2)	100(0)	100(0)	100(0)
	41.0	95(3.4)	100(0)	99(3.8)	100(0)	86(11.5)	86(13.5)	96(5.9)	95(10.1)	99(2.7)
	23	90(3.6)	96(6.1)	95(7.8)	99(1.7)	67(17.1)	92(9.0)	83(11.5)	93(8.2)	90(12.1)
6	128	97(1.9)	95(6.2)	99(2.8)	98(5.1)	96(7.1)	94(6.3)	97(5.5)	99(3.8)	100(0)
	41.4	96(2.5)	98(5.3)	90(11.4)	96(5.9)	99(2.9)	96(7.7)	99(3.6)	93(9.4)	100(0)
	26	93(2.7)	91(7.4)	94(6.9)	95(6.2)	97(5.4)	96(6.0)	84(11.2)	93(10.8)	96(5.9)
7	128	98(1.7)	98(5.1)	95(7.4)	93(6.2)	99(2.8)	100(0)	97(5.5)	100(0)	100(0)
	38.7	95(2.4)	99 (1.2)	95(6.1)	89(10.9)	90(10.0)	98(5.1)	100(0)	97(5.7)	92(10.7)
	27	95(2.4)	97(5.2)	94(7.2)	84(12.3)	95(6.2)	91(9.2)	97(6.1)	100(0)	100(0)

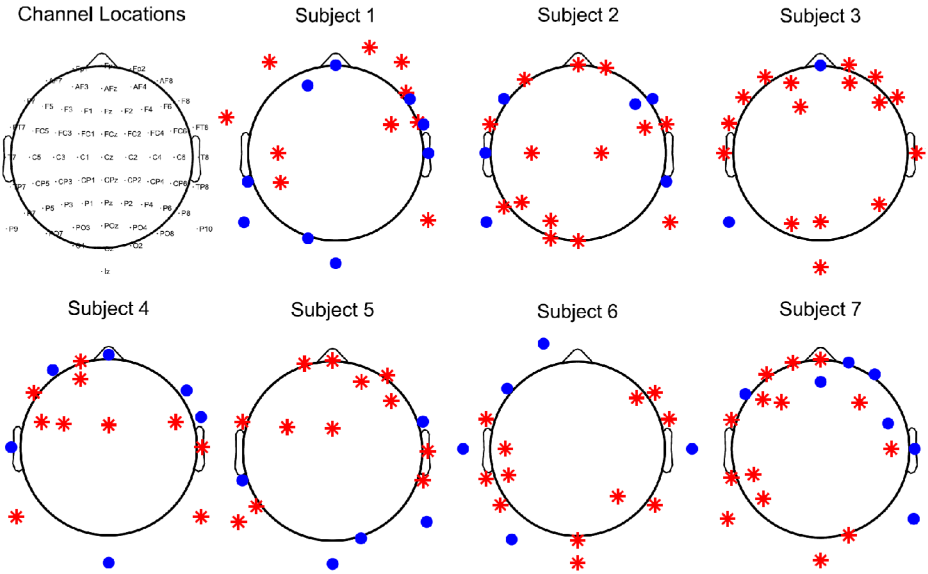

Fig. 1. Plot of significant channels for all subjects in the study. The first plot depicts the 10-20 channel orientation of a 64-channel Biosemi System. Channels with red stars indicate that at least one of the two AR(2) coefficients was significant in at least 75% of bootstrap samples, while channels with blue circles indicate both AR(2) coefficients were significant at this same criterion.

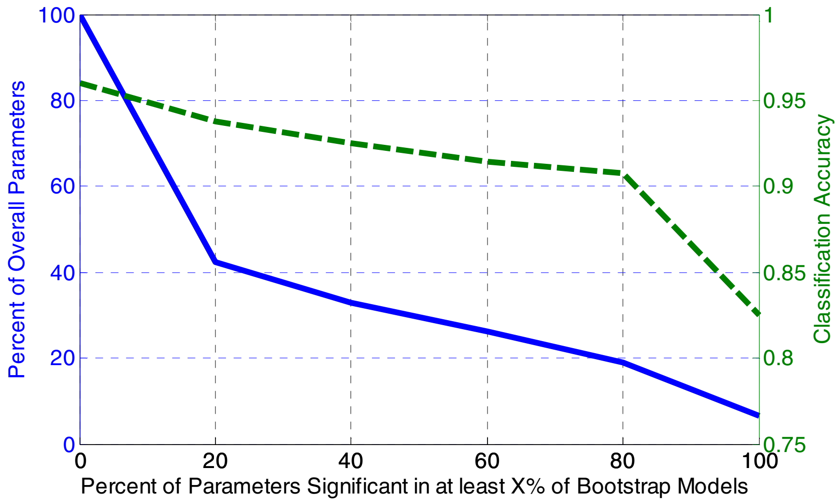

Fig. 2. Plot of the average percentages of overall parameters and the classification percentage for different percentage of parameters observed in bootstrap models. The dashed green line denotes the classification accuracy, while the solid blue line denotes the percent of overall parameters used in the model.

5 Conclusion

In this paper we have proposed a method for down-selecting the appropriate features necessary for accurate discrimination of EEG artifacts based on elastic net penalized regression models. The elastic net penalty applied to multinomial regression can effectively handle the high correlations and redundancy in the AR parameters and appears to be an effective general approach for feature selection in EEG analysis. In our analysis, using the elastic net penalty with multinomial regression effectively reduced the number of parameters by 60% without any loss in classification accuracy. The overall classification accuracy remained above 90% until we restricted the number of parameters to less than 20% of the overall parameters available (Fig 2). This indicates that a significant computational savings could be achievable if implemented in a BCI system. For example, data streamlining is critical in new wireless EEG headsets, where transmission bandwidth is limited by power. Although the high variability observed across subjects might limit the possibility of physically tailoring the channel locations to a specific user, one possible scheme might be to only record and broadcast data from the channels previously established to be most meaningful for that individual. Potential applications of this approach include monitoring subjects for artifact instances such as eye blink frequency and duration for detecting lapses in attention during experiments [16].

The results derived from artifact classification by the regularized multinomial regression are corroborated by known brain physiological properties. For example, there were many frontal channels identified as being highly significant, which is expected given that these channels exhibit eye movement artifacts the most strongly. Meanwhile, there were also many significant channels located around the edges of the cap, while the majority of those in the center are less likely to significantly contribute to the discrimination. One possible reason for this is that muscle activations from the rotate head (RH) condition are picked up by the channels located near the neck. Channels near the ears are also significant in many subjects, as these channels are located near the jawline and pick up jaw clench and jaw movement artifacts. A few channels located at the top are most likely contributing to the model of the baseline condition, as these channels are minimally impacted by artifacts.

Acknowledgments. We thank Scott Kerick and Kaleb McDowell of the Army Research Laboratory for helpful discussions and for help with data collection. This research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0022. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

1. Lin, C.-T., Chang, C.-J., Lin, B.-S., Hung, S.-H., Chao, C.-F., Wang, I.-J.: A Real-Time Wireless Brain-Computer Interface System for Drowsiness Detection. IEEE Transactions on Biomedical Circuits and Systems 4, 214–222 (2010)
2. Lance, B.J., Kerick, S.E., Ries, A.J., Oie, K.S., McDowell, K.: Brain-Computer Interface Technologies in the Coming Decades. Proceedings of the IEEE 100, 1585–1599 (2012)
3. Baccalá, L.A., Sameshima, K.: Partial directed coherence: a new concept in neural structure determination. Biol. Cybern. 84, 463–474 (2001)
4. Möller, E., Schack, B., Arnold, M., Witte, H.: Instantaneous multivariate EEG coherence analysis by means of adaptive high-dimensional autoregressive models. J. Neurosci. Methods 105, 143–158 (2001)
5. Franaszczuk, P.J., Bergey, G.K., Kamiński, M.J.: Analysis of mesial temporal seizure onset and propagation using the directed transfer function method. Electroencephalography and Clinical Neurophysiology 91, 413–427 (1994)
6. Anderson, C.W., Stolz, E.A., Shamsunder, S.: Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks. IEEE Transactions on Biomedical Engineering 45, 277–286 (1998)
7. Ge, D., Srinivasan, N., Krishnan, S.M.: Cardiac arrhythmia classification using autoregressive modeling. BioMedical Engineering OnLine 1, 5 (2002)
8. Übeyli, E.D.: Least squares support vector machine employing model-based methods coefficients for analysis of EEG signals. Expert Systems with Applications 37, 233–239 (2010)
9. Van de Velde, M., Ghosh, I.R., Cluitmans, P.J.M.: Context related artefact detection in prolonged EEG recordings. Computer Methods and Programs in Biomedicine 60, 183–196 (1999)
10. Lawhern, V., Hairston, W.D., McDowell, K., Westerfield, M., Robbins, K.: Detection and classification of subject-generated artifacts in EEG signals using autoregressive models. J. Neurosci. Methods 208, 181–189 (2012)
11. Friedman, J., Hastie, T., Tibshirani, R.: Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software 33, 1 (2010)
12. Zhu, J., Hastie, T.: Classification of gene microarrays by penalized logistic regression. Biostatistics 5, 427–443 (2004)
13. U.S. Department of Defense, Office of the Secretary of Defense: Code of federal regulations, protection of human subjects. Government Printing Office. 32 CFR 19 (1999)
14. U.S. Department of the Army: Use of volunteers as subjects of research. Government Printing Office. AR 70-25 (1990)
15. Schlögl, A.: A comparison of multivariate autoregressive estimators. Signal Processing 86, 2426–2429 (2006)
16. Kim, Y.S., Baek, H.J., Kim, J.S., Lee, H.B., Choi, J.M., Park, K.S.: Helmet-based physiological signal monitoring system. Eur. J. Appl. Physiol. 105, 365–372 (2009)